

Introduction

Telecom Customer Churn

by Konstantinos Tzortzakis

Customer churn occurs when customers or subscribers stop doing business with a company or service, also known as customer attrition. It is also referred to as loss of clients or customers. One particular industry in which churn rates are particularly useful is the telecommunications industry, simply because most customers have multiple options from which to choose within a geographic location.

The dataset I am going to use for this project is " **WA_Fn UseC_ Telco Customer Churn.csv** " provided by ibm.com

Source: https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_Fn-UseC_-Telco-Customer-Churn.csv?cm_mc_uid=42350349240115511137586&cm_mc_sid_50200000=48695861551226681996&cm_mc_sid_52640000=68226791551226682000

This dataset comes from the 7,043 customers' data of a telecommunications company in order to predict behavior to retain customers. Although, further information about the source of the data is not given (when and how the data was collected), we know that the data is updated on a monthly basis because we are aware of the customers that left this company within the last month. The output of this project will be to classify the customers in two categories according to the likelihood of them leaving this telecom company in the future. It is very important for a company (especially for a highly competitive telecom company) to understand their customers' behavior, especially the ones who left and the reason behind this decision in order to provide incentives (action) like service discounts and more attractive contracts to the other customers to maintain this relationship (outcome). As a result, the company will be able to sustain its revenue or even grow it, which is the ultimate goal of every company.

To achieve that I have to work with a data of 7043 customers (7043 rows) and 21 features (21 columns). My output is the "Churn" column. The features:

- customer id

Customers' demographics:

- gender (male, female)
- SeniorCitizen (0,1) , change "0 or 1" to "yes or no" as the rest features
- Partner (yes, no)
- Dependents (yes, no)

Services that each customer has signed up for:

- PhoneService (yes, no)
- MultipleLines (yes, no, no phone service), one easy fix to clear that is to change “no service” to “no”.
- InternetService (Fiber optic, DSL, no)
- OnlineSecurity (yes, no, no internet service)
- OnlineBackup (yes, no, no internet service)
- DeviceProtection (yes, no, no internet service)
- TechSupport (yes, no, no internet service)
- StreamingTV (yes, no, no internet service)
- StreamingMovies (yes, no, no internet service), my initial thought is to change “no internet service” to “no” for these 7 features

Customer account information:

- Contract (one year, two year, month to month)
 - PaperlessBilling (yes, no)
 - PaymentMethod (electronic check, mailed check, bank transfer(automatic), credit card(automatic))
 - MonthlyCharges (number) in \$
 - TotalCharges (number) in \$
 - Tenure (min: 1, max:72) in months
-
- Churn (yes, no): customers who left within the last month

Variables such as Contract, Tenure, and Total Charges are very important factors. For example, a customer with month to month contract is far more likely to leave the company rather than a customer with two-year contract.

This data has been used in the past in Kaggle from multiple users who tried to do the best exploratory data analysis (using python or R) and find the best prediction (using Logistic Regression, Random Forest, ANN). Given that this is a popular dataset, I will do my own analysis and then merge it with an external dataset regarding customer expenditure from the Bureau of Labor Statistics:

<https://www.bls.gov/cex/tables.htm>

I will make the assumption that these telecom customers belong to the sample of the survey in order to be able to expand my questions and come up with more compelling and unique solutions.

Personally, I am really interested to work with this kind of data because customer churn is a huge issue that can be monitored especially nowadays that subscriptions are becoming one of the most popular business models (for example Netflix, Amazon, Apple Music, Spotify). Furthermore, my sprint project in the MSIM program

which involves a consumer product, is also considering using the subscription model which motivates me even more to become an expert in this particular field.

Framework:

